frontiers

Editorial: "ViTac: Integrating Vision and Touch for Multimodal and Cross-Modal Perception"

Shan Luo [1,*], **Nathan F. Lepora** [2], **Uriel Martinez-Hernandez** [3], **Joao Bimbo** [4] **and Huaping Liu** [5]

[1] *smARTLab, Department of Computer Science, University of Liverpool, Liverpool, United Kingdom, E-mail: shan.luo@liverpool.ac.uk*
[2] *Department of Engineering Mathematics and Bristol Robotics Laboratory, University of Bristol, Bristol, United Kingdom, E-mail: n.lepora@bristol.ac.uk.*
[3] *inte-R-action Lab and Centre for Autonomous Robotics (CENTAUR), University of Bath, Bath, United Kingdom, E-mail: u.martinez@bath.ac.uk*
[4] *Yale University, United States, E-mail: joao.bimbo@yale.edu*
[5] *Tsinghua University, China, E-mail: hpliu@tsinghua.edu.cn*

Correspondence*:
shan.luo@liverpool.ac.uk

1 INTRODUCTION

Animals interact with the world through multimodal sensing inputs, especially vision and touch sensing in the case of humans interacting with our physical surroundings. In contrast, artificial systems usually rely on a single sensing modality, with distinct hardware and algorithmic approaches developed for each modality. For example, computer vision and tactile robotics are usually treated as distinct disciplines, with specialist knowledge required to make progress in each research field. Future robots, as embodied agents interacting with complex environments, should make best use of all available sensing modalities to perform their tasks.

Over the last few years, there have been advances in the fusing of information from distinct modalities and selecting between those modalities to use the most appropriate information for achieving a goal; for example grasping or manipulating objects in a desired manner, such as stacking objects in storage crates or folding clothing. We have seen a shift in the ways of linking vision and touch, from combining hand-crafted features (Luo et al. (2015, 2017)) to learning a shared latent space with deep neural networks (Luo et al. (2018); Lee et al. (2019b)). The integration of the visual and haptic cues have been shown to enhance perception in one modality, either enabling better tactile understanding of haptic adjectives (Gao et al. (2016)) or learning visual representations (Pinto et al. (2016)), and also result in better performance in achieving a task (Luo et al. (2018); Lee et al. (2019b)).

Another trend is that there has been a recent acceleration in the development of optical tactile sensors using cameras, such as the GelSight (Yuan et al. (2017a); Johnson and Adelson (2009)) and TacTip (Ward-Cherrier et al. (2018); Chorley et al. (2009)), which bridge the gap between vision and tactile sensing to create cross-modal perception. On the one hand, this crossover has enabled techniques developed for computer vision to be applied to tactile sensing; examples include the use of convolutional neural networks for estimating contact variables directly from the tactile images (Yuan et al. (2017b); Lepora et al. (2019)), and also sim-to-real methods that were pioneered in robot vision but are now finding application in robot touch (Fernandes et al. (2021)). On the other hand, there have been the development of methods that connect the look and feel of objects being interacted with (Calandra et al. (2017)), progressing more

recently to methods that can transform between or match visual and tactile data (Takahashi and Tan (2019); Lee et al. (2019a); Li et al. (2019)).

2 CONTENTS OF THE RESEARCH TOPIC

The contents of the Research Topic include four papers on topics addressing diverse challenges of multimodal and cross-modal perception with vision and touch.

In "A Framework for Sensorimotor Cross-Perception and Cross-Behavior Knowledge Transfer for Object Categorization", Tatiya *et al.* propose a framework for knowledge transfer across exploratory behaviors, e.g., press, grasp and hold, and sensory modalities, with audio, haptic, vibrotactile and visual feedback used. They use two models based on variational auto-encoders and encoder-decoder networks respectively to map shared multi-sensory object observations of a set of objects across different robots. Their results of categorising 100 objects indicate that sensorimotor knowledge about objects can be transferred both across behaviors and across sensory modalities, which can boost the robot's capability in cross-modal and cross-behavior perception.

Haptic information can also be conveyed to a user during teleoperation tasks in the form of visual cues. This cross-modal sensation between visual and tactile sensing is usually termed pseudo-haptics (Li et al. (2016)). In "Proposal and Evaluation of Visual Haptics for Manipulation of Remote Machine System" by Haruna *et al.*, this approach is assessed using electroencephalogram (EEG) data on a Virtual Reality (VR) setting. The authors carried out a human-subject experiment where users were asked to perform a series of pick-and-place tasks with and without displaying a visual overlay with haptic information. The main finding of this paper is that the added pseudo-haptic information improves the usability of teleoperation systems and reduces users' cognitive load. Additionally, the results also suggest that the brainwave information flow can be used as a quantitative measurement of a system's usability or "familiarity".

In "Using Tactile Sensing to Improve the Sample Efficiency and Performance of Deep Deterministic Policy Gradients (DDPG) for Simulated In-Hand Manipulation Tasks", Andrew Melnik and co-authors demonstrate the importance of tactile sensing for manipulation with an anthropomorphic robot hand. They perform in-hand manipulations tasks such as reorienting a held block, using a simulation model of the Shadow Dexterous Robot Hand covered with 92 virtual touch sensors. Deep reinforcement learning methods such as DDPG are known to require a large number of training samples that can make them impractical to train in physical environments. The authors showed that tactile sensing data can improve sample efficiency up to three-fold with a performance gain of up to 46%, on several simulated manipulation tasks.

GelTip is an optical tactile sensor proposed by Gomes *et al.* in the work entitled "Blocks World of Touch: Exploiting the Advantages of All-around Finger Sensing in Robot Grasping". This sensor is composed of a rounded fingertip fully covered with an elastomer and with a camera mounted at the sensor base. The deformations of the elastomer that occur when the fingertip touches an object are captured and precisely tracked by the camera. The geometry of the sensor and location of the camera allows the device to detect deformations from any position on the fingertip, which contrasts with the reduced contact region found in traditional optical sensors. GelTip can identify contact location processes with 1mm precision. This sensor has been mounted on a robotic gripper and successfully tested with object touch and grasping tasks. This process has shown that the sensor is capable of detecting collisions from any orientation while approaching to an object, which can be used by the robotic gripper to the grasping strategy. Overall, GelTip offers an effective optical tactile sensor with the potential to enable many robotic gripping and manipulation tasks that require tactile sensing from all over the fingertip.

REFERENCES

Calandra, R., Owens, A., Upadhyaya, M., Yuan, W., Lin, J., Adelson, E. H., et al. (2017). The feeling of success: Does touch sensing help predict grasp outcomes? In *Conference on Robot Learning*. 314–323

Chorley, C., Melhuish, C., Pipe, T., and Rossiter, J. (2009). Development of a tactile sensor based on biologically inspired edge encoding. In *IEEE International Conference on Advanced Robotics*

Fernandes, D. G., Paoletti, P., and Luo, S. (2021). Generation of GelSight Tactile Images for Sim2Real Learning. *IEEE Robotics and Automation Letters*

Gao, Y., Hendricks, L. A., Kuchenbecker, K. J., and Darrell, T. (2016). Deep learning for tactile understanding from visual and haptic data. In *IEEE International Conference on Robotics and Automation (ICRA)*. 536–543

Johnson, M. K. and Adelson, E. H. (2009). Retrographic sensing for the measurement of surface texture and shape. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1070–1077

Lee, J.-T., Bollegala, D., and Luo, S. (2019a). "Touching to See" and "Seeing to Feel": Robotic Cross-modal Sensory Data Generation for Visual-Tactile Perception. In *IEEE International Conference on Robotics and Automation (ICRA)*. 4276–4282

Lee, M. A., Zhu, Y., Srinivasan, K., Shah, P., Savarese, S., Fei-Fei, L., et al. (2019b). Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. In *IEEE International Conference on Robotics and Automation (ICRA)*. 8943–8950

Lepora, N. F., Church, A., De Kerckhove, C., Hadsell, R., and Lloyd, J. (2019). From pixels to percepts: Highly robust edge perception and contour following using deep learning and an optical biomimetic tactile sensor. *IEEE Robotics and Automation Letters* 4, 2101–2107

Li, M., Sareh, S., Xu, G., Ridzuan, M. B., Luo, S., Xie, J., et al. (2016). Evaluation of pseudo-haptic interactions with soft objects in virtual environments. *PloS One* 11, e0157681

Li, Y., Zhu, J.-Y., Tedrake, R., and Torralba, A. (2019). Connecting touch and vision via cross-modal prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10609–10618

Luo, S., Bimbo, J., Dahiya, R., and Liu, H. (2017). Robotic tactile perception of object properties: A review. *Mechatronics* 48, 54–67

Luo, S., Mou, W., Althoefer, K., and Liu, H. (2015). Localizing the object contact through matching tactile features with visual map. In *IEEE International Conference on Robotics and Automation (ICRA)*. 3903–3908

Luo, S., Yuan, W., Adelson, E., Cohn, A. G., and Fuentes, R. (2018). ViTac: Feature sharing between vision and tactile sensing for cloth texture recognition. In *IEEE International Conference on Robotics and Automation (ICRA)*. 2722–2727

Pinto, L., Gandhi, D., Han, Y., Park, Y.-L., and Gupta, A. (2016). The curious robot: Learning visual representations via physical interactions. In *European Conference on Computer Vision*. 3–18

Takahashi, K. and Tan, J. (2019). Deep visuo-tactile learning: Estimation of tactile properties from images. In *IEEE International Conference on Robotics and Automation (ICRA)*. 8951–8957

Ward-Cherrier, B., Pestell, N., Cramphorn, L., Winstone, B., Giannaccini, M. E., Rossiter, J., et al. (2018). The TacTip family: Soft Optical Tactile Sensors with 3D-printed Biomimetic Morphologies. *Soft Robotics* 5, 216–227

Yuan, W., Dong, S., and Adelson, E. H. (2017a). Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors* 17, 2762

Yuan, W., Zhu, C., Owens, A., Srinivasan, M. A., and Adelson, E. H. (2017b). Shape-independent hardness estimation using deep learning and a gelsight tactile sensor. In *IEEE International Conference on Robotics and Automation (ICRA)*. 951–958